



## Discours

Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics

7 | 2010  
Varia

---

# The "Fame Effect" or How the syntactic choices of writers can be explained by their assumptions about their addressees' state of knowledge: the case of relevance-oriented, non-restrictive noun modifiers

Rudy Loock

---



### Electronic version

URL: <http://journals.openedition.org/discours/8027>

DOI: 10.4000/discours.8027

ISSN: 1963-1723

### Publisher:

Laboratoire LATTICE, Presses universitaires de Caen

### Electronic reference

Rudy Loock, « The "Fame Effect" or How the syntactic choices of writers can be explained by their assumptions about their addressees' state of knowledge: the case of relevance-oriented, non-restrictive noun modifiers », *Discours* [Online], 7 | 2010, Online since 20 December 2010, connection on 19 April 2019. URL : <http://journals.openedition.org/discours/8027> ; DOI : 10.4000/discours.8027

---



*Discours* est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.



## **The “Fame Effect” or How the syntactic choices of writers can be explained by their assumptions about their addressees’ state of knowledge: the case of relevance-oriented, non-restrictive noun modifiers**

---

Rudy Loock

Université de Lille 3 & UMR STL 8163 du CNRS

.....  
Rudy Loock, « The “Fame Effect” or How the syntactic choices of writers can be explained by their assumptions about their addressees’ state of knowledge: the case of relevance-oriented, non-restrictive noun modifiers », *Discours*, 7 | 2010, [En ligne], mis en ligne le 20 décembre 2010.

.....  
URL : <http://discours.revues.org/8027>. Consulté le 20 décembre 2010.



# The “Fame Effect” or How the syntactic choices of writers can be explained by their assumptions about their addressees’ state of knowledge: the case of relevance-oriented, non-restrictive noun modifiers <sup>1</sup>

---

Rudy Loock

Université de Lille 3 & UMR STL 8163 du CNRS

.....

In this paper we would like to explore, within Vallduví’s (1992, 1993) information packaging theory, how writers’ assumptions about their addressees’ state of knowledge can account for the choice writers make between competing syntactic structures, which we call “allostructures”, to convey the same informational content. The assumed old/new informational status has then an influence on the grammatical choices made by the writers. In particular we investigate here, using the web as corpus, the constraints that govern writers’ choice between the following non-restrictive modifiers of nouns that aim at optimizing the relevance (as defined by Sperber & Wilson) of the referents denoted by the nouns they modify: nominal appositives, appositive relative clauses, non-restrictive pre-modifiers.

**Keywords:** informational status, familiarity, relevance, appositive relative clauses, nominal appositives, allostructures, information packaging, web as corpus

.....

*Cet article explore la façon dont les hypothèses de l’énonciateur sur les connaissances de son / ses destinataire(s) peuvent avoir une influence sur ses choix entre des structures syntaxiques concurrentes (ou « allostructures ») afin de véhiculer un contenu informationnel donné. Le statut informationnel ancien / nouveau supposé aura alors une influence sur les choix grammaticaux effectués par celui-ci. Dans le cadre de la théorie de l’information packaging telle que développée par Vallduví (1992, 1993) et à partir d’une étude de corpus qui utilise l’internet comme source principale, nous étudions en particulier les contraintes qui régissent le choix entre des modificateurs non restrictifs à visée de pertinence (au sens de Sperber et Wilson), à savoir les propositions subordonnées relatives appositives, les appositions nominales, et les prémodificateurs.*

**Mots clés:** statut informationnel, familiarité, pertinence, relatives appositives, apposition nominale, allostructures, information packaging, internet comme corpus

## Introduction

1 At the risk of stating the obvious, there is no discourse without a speaker/writer and his/her intended addressees. The identity of the former and the abstract conception that s/he has about the latter and the state of their knowledge play an essential role in the lexical but also grammatical choices that are made. Like the co-text, the

---

1. We would like to thank the audience of the Caen workshop “What Texts Do to Sentences” that was held in December 2009 for their useful comments. We are also grateful to the anonymous reviewers of this journal for providing helpful suggestions. Finally, we thank Kathleen M. O’Connor for her input on a previous draft of this article. Any remaining errors are naturally our own.

situational context (identity of speaker/writer and addressees but also time and place of utterance) has therefore an influence on the acceptability of sentences.

- 2 The aim of this paper is to try and establish the existence of what we call a “fame effect”, which could explain writers’ choices between different syntactic structures to convey the same informational content/logico-semantic meaning. Within Vallduví’s information packaging framework, we investigate the influence of the hearer new/hearer old informational status on the possibility to use relevance-oriented, non-restrictive modifiers such as appositive relative clauses, nominal appositives, pre-modifiers etc. According to our hypothesis, the hearer old status of an identity relationship A is B has an influence on the structures that can be used by writers. Fame is therefore to be understood here not only as actual fame (“the state of being known about by a lot of people because of your achievements”, *Longman Dictionary of Contemporary English*), but specifically as referring to familiarity: are the referent and its identity/role/function known by the addressee(s) or not?

- 3 The paper is organized as follows. In the first section we expose and define our starting point, that is relevance appositive relative clauses (henceforth ARCs) such as defined in Loock’s (2003, 2005, 2007, 2010) taxonomy of ARCs’ discourse functions, as well as their competing structures. In the second section we list the constraints that govern writers’ choices between ARCs and their competing structures, with special emphasis on what we call the “fame effect”, to which this paper is dedicated. Section 3 investigates the way such a “fame effect” can be defined and verified, with particular emphasis on methodological issues such as why and how the web can be used as a corpus in spite of its limitations.

## 1. Starting point: relevance appositive relative clauses and their competing structures

### 1.1. A taxonomy of appositive relative clauses’ discourse functions

- 4 In Loock (2003, 2005, 2007, 2010), we have suggested the existence of three main discourse functions for appositive relative clauses (ARCs) in English (also called non-restrictive or non-defining), filling in a blank in the vast literature on the subject. We have defined three main discourse functions: continuative ARCs [1], which support the trajectory of the narrative by showing the events in a sequence with a possible causal link; relevance ARCs [2], whose aim is to optimize the relevance of the antecedent and/or the subject-predicate relation within the main clause; subjectivity ARCs [3], which convey information that is explicitly subjective and results in a disjunction with the main clause<sup>2</sup>.

2. The source of examples is specified in brackets. When no source is indicated, this means the example has been invented or is an attested example that has undergone modification for the purposes of our demonstration. The symbols used to show the degree of acceptability of the utterances are: “\*” for a completely ungrammatical sentence, “#” for a pragmatically unacceptable utterance, “?” for questionable

- [1] So we asked a man, who shrugged his shoulders and disappeared into a nearby shop.  
(*A View of Wales*, tourist brochure)
- [2] It is hard to square his action with the Energy Policy and Conservation Act, which authorizes the president to tap the reserve.  
(*International Herald Tribune* 36, 576, p. 6)
- [3] This incredible spirit – which Chelsea so clearly lack – is summed up beautifully by Gemmill, who has been unable to command a regular plane and has also been a target for some of the fans on his rare appearances.  
(*The Mirror*, 11/27/00)

## 1.2. Relevance appositive relative clauses

5 In this paper, we would like to focus on relevance ARCs such as [2] or [4]:

- [4] Tony Sewell, who has just finished an inquiry into soaring levels of exclusions among black pupils from a London school, claimed that too much concern with money and consumer goods was almost as damaging to black pupils' chances as racism.  
(*Guardian Weekly* 163, 9)

6 Such ARCs provide non-restrictive information on the antecedents *the Energy Policy and Conservation Act* and *Tony Sewell* which, in spite of their referential stability, are not sufficiently 'determined', for at least some of the addressees, to be used alone in discourse. The aim of the relevance ARCs is thus to optimize the relevance of these antecedents within the discourse, in line with Sperber & Wilson's definition of the principle of relevance (Sperber & Wilson, 1986: 125, 158 ff.):

Extent condition 1: an assumption is relevant in a context to the extent that its contextual effects in this context are large.

Extent condition 2: an assumption is relevant in a context to the extent that the effort required to process it in this context is small.

7 The principle of relevance thus corresponds to the search for adequate contextual effects for no gratuitous processing effort, a contextual effect consisting of a modification of what Sperber & Wilson call the "background", *i. e.* the context of hypotheses known to the addressee(s), while the mental effort corresponds to the amount of effort necessary for the addressee(s) to process these contextual effects.

8 The use of relevance ARCs as we define them is thus triggered by the speaker/writer's will to optimize the relevance of his/her utterances by reducing the mental effort on the part of the addressee(s) and at the same time increasing the contextual effects. Speakers/writers then take precautions with respect to the knowledge they assume is shared with the addressee(s) (hearer old information). The use of a

---

pragmatic acceptability, '??' for very questionable pragmatic acceptability. The judgments were performed by the author of the article himself, as well as by a native speaker of English.

relevance ARC is thus triggered by the speaker/writer's desire to avoid any violation of the relevance principle as defined by Sperber & Wilson. In this article, we study relevance ARCs alongside other relevance-oriented structures (see the next sub-section); that is, structures that convey non-restrictive information aimed at optimizing the relevance of a specific referent and, by doing so, the relevance of the subject-predicate relation in the main clause as a whole; the informational status of the referent and the informational content of such relevance-oriented structures is crucially indeterminable (hearer new or hearer old depending on the addressees' knowledge store). It is this last aspect (hearer new or old informational status, in other words the degree of familiarity) that this article focuses on, alongside its influence on grammatical choices.

### 1.3. Competing structures

- 9 Naturally, ARCs are not the only syntactic structures that fulfill such a relevance-oriented function. Among other non-restrictive modifiers, we find nominal appositives, non-restrictive pre-modifiers, sentential parentheticals, independent clauses, co-referential NPs via unfaithful anaphora (as defined by Blanche-Benveniste & Chervel, 1966). Below we provide the same informational content in sentences using these different competing structures, starting with the original example, taken from the web:

- [5] A. Raja, the Indian environment minister, said his country would accept help to reduce emissions but would not be forced into cuts. (NOMINAL APPOSITIVE)  
("India ignores Kyoto demands", Reuters, 01/12/06)
- [6] A. Raja, who is the Indian environment minister, said his country would accept help to reduce emissions but would not be forced into cuts. (APPOSITIVE RELATIVE CLAUSE)
- [7] Indian environment minister A. Raja said his country would accept help to reduce emissions but would not be forced into cuts. (NON-RESTRICTIVE PRE-MODIFIER)
- [8] A. Raja – he is the Indian environment minister – said his country would accept help to reduce emissions but would not be forced into cuts. (SENTENTIAL PARENTHETICAL)
- [9] Speaking after the first meeting of a climate change group created by six of the world's biggest polluters in Sydney, A. Raja<sub>i</sub> said his country would accept help to reduce emissions but would not be forced into cuts. "Neither the Kyoto Protocol nor this partnership can stipulate anything upon the government of India to reduce emissions," the Indian environment minister<sub>i</sub> said. (CO-REFERENTIAL NP)
- [10] A. Raja said his country would accept help to reduce emissions but would not be forced into cuts. He is the Indian environment minister. (INDEPENDENT CLAUSE)



10 Sentences [5]–[10] provide the same informational content, the same logico-semantic meaning: (i) *A. Raja said X* and (ii) *A. Raja is the Indian environment minister* (underlined elements). This second piece of information is non-restrictive (there is only one A. Raja) and relevance-oriented: the speaker/writer provides supplementary information about A. Raja that might be known (hearer old information) or not (hearer new information) by the addressees. By inserting such information, the speaker/writer ensures the relevance of the discourse as a whole (not knowing who A. Raja is reduces the contextual effects such as defined by Sperber & Wilson).

11 If we place ourselves within the information packaging theory such as defined by Vallduví (1992, 1993), constraints exist which govern speakers'/writers' choices between such competing syntactic structures. Information packaging was first defined by Prince as "the tailoring of an utterance by a sender to meet the particular assumed needs of the intended receiver" (Prince, 1981: 224). The speaker's/writer's choice of a specific packaging then "reflects the sender's hypotheses about the receiver's assumptions and beliefs and strategies". As a consequence, just as allophones represent the different phonetic realizations of the same phoneme or allomorphs the different realizations of the same morpheme, allostructures represent the different possible syntactic realizations for the same informational content (Loock, 2005, 2010)<sup>3</sup>. Sentences [5]–[10] show non-restrictive, relevance-oriented allostructures. The choice between these competing syntactic structures or allostructures represents an instruction (Vallduví, 1992, 1993) that tells the addressee(s) how the information is to be received and processed; the meaning of the whole sentence consists in the addition of the propositional content and the instruction provided by the choice of structure. Within such a theoretical framework, sentences [5]–[10] are not interchangeable in discourse and constraints must govern the choice between allostructures. Felicitousness is not random but dependent on the co-text and context that represent such constraints. In the next section we define and illustrate such constraints, with special emphasis on what we call the familiarity constraint or "fame effect".

## 2. The definition of constraints

12 Syntactic-related phenomena such as end-weight phenomena and syntactic length/complexity first need to be teased apart from pragmatics-related phenomena. In such cases constraints that govern speakers'/writers' choices have nothing to do with the semantics and/or pragmatics of the utterance within its context, as speakers/writers use a structure that fits in the syntactic organization of the discourse. Example [11], for instance, shows the use of an independent clause conveying supplementary information about *Bolland* (hearer new for at least some of the addressees, probably

3. This term was coined after the term *allosentences*, which was first introduced by Daneš (1966) and Chafe (1976), and specifically defined by Lambrecht (1996: 35) as "semantically equivalent but formally and pragmatically divergent sentence pairs".

for most of them) to ensure the relevance of the discourse as a whole. A relevance ARC [11a] cannot be inserted because of the use of the genitive construction<sup>4</sup>:

[11] [Prince] Charles was “amazed” and “shocked” by Bolland’s comments yesterday to a Sunday newspaper. He is the prince’s former deputy private secretary and press adviser.  
(*The Mirror*, 01/09/04)

[11a] \*Charles was “amazed” and “shocked” by Bolland, who is the prince’s former deputy private secretary and press advisers,’s comments yesterday to a Sunday newspaper.

13 Other types of constraints, related to semantic and pragmatic phenomena are, among others, (i) (one of) class membership *vs* identity relation (Burton-Roberts, 1975) or (ii) the specificational *vs* identificational reading (Higgins, 1979) of an ‘A is B’ identity relation.

14 Following Burton-Roberts’s (1975) well-known distinction, the existence of a class membership relation between the two units of a nominal appositive construction generally makes the reformulation by an ARC systematically possible ([12]–[12a]), while the reformulation is most of the time not possible with an identity relation ([13]–[13a]):

[12] “I think people were scared,” said Peter Prows, a politics student form Oberlin College, Ohio.  
(*The Guardian Europe*, 10/01/01)

[12a] “I think people were scared,” said Peter Prows, who is a politics student form Oberlin College, Ohio.

[13] The King of Pop, Michael Jackson, has died today.  
([http://regator.com/p/208471659/king\\_of\\_pop\\_michael\\_jackson\\_dead\\_at\\_50/](http://regator.com/p/208471659/king_of_pop_michael_jackson_dead_at_50/))

[13a] ??/#The King of Pop, who is/was Michael Jackson, has died today.

15 One exception to this is the existence of a specificational reading as defined by Higgins (1979) for the identity relation [13b]<sup>5</sup>:

4. In the newspaper article from which this sequence has been extracted, devoted to some comments made by Bolland about Princess Diana, the use of a periphrastic genitive (*the comments of Bolland*) is not felicitous.

5. Higgins (1979) has defined two types of reading for copular clauses (one of class membership). Consider [i], taken from Mikkelsen (2004):

[i] The doctor on call, Dr Jones, is to arrive at 9 p.m.

The first reading corresponds to a situation where two nurses know Dr. Jones and work in his unit, in which case the apposed unit does not identify the doctor but specifies who the doctor on call is among n other possible doctors (The doctor on call is Dr Jones, not Dr Smith or Dr Wilcockson for instance). The second reading corresponds to a situation where a nurse for instance provides the name of the doctor on call to a patient. For more details see Loock (2010: 170–173).

- [13b] The King of Pop, who is Michael Jackson, has died today. (and not Prince or David Bowie, for instance)

16 Focusing on identity relations as defined above<sup>6</sup>, what we would like to investigate in this paper, however, is another kind of pragmatics-related constraint, *viz* a familiarity constraint. In other words, we would like to check how writers' assumptions about their addressees' state of knowledge can account for the choice writers make between the non-restrictive allostructures illustrated in [5]–[10] to convey the same informational content. It seems that there is a link between the assumed familiarity (*i. e.* hearer old informational status) and the use of some of the relevance-oriented syntactic structures. The writer's assumptions about the addressees' knowledge then have an influence on the choices made by the writer between *n* possible syntactic vehicles. In this paper we focus mainly on ARCs, nominal appositives and pre-modifiers.

17 Consider the following examples, which show the felicitous use of nominal appositives establishing an identity relation between two elements:

- [14] Some expected Barack Obama, the president of the United States, to appoint a completely new economic team so as to implement another New Deal.  
(adapted from <http://www.globalresearch.ca/index.php?context=va&aid=11253>)
- [15] Bill Clinton, the former president of the United States, will attend an international seminar on AIDS and SARS (severe acute respiratory syndrome) and deliver a lecture on global AIDS prevention and control efforts, a seminar official said Friday.  
(adapted from [http://english.peopledaily.com.cn/200311/08/eng20031108\\_127857.shtml](http://english.peopledaily.com.cn/200311/08/eng20031108_127857.shtml))
- [16] Angela Merkel, the German chancellor, on Friday described Barack Obama's presidency as a "unique opportunity" to revive the Middle East peace process as the US leader continued his international tour with a stop in the historic eastern city of Dresden.  
(example from the Internet, no longer available)
- [17] Nancy Pelosi, the Speaker of the House [...], is among those on the Left now seeking to find common ground with the conservative populism that is sweeping across the United States.  
(<http://www.telegraph.co.uk/news/worldnews/northamerica/usa/barackobama/7354180/Democrats-including-Nancy-Pelosi-jump-on-Tea-Party-bandwagon.html>)
- [18] According to Arne Duncan, the secretary of education, the president will discuss the importance of hard work, educational goals and other topics.  
(<http://www.myfox8.com/topic/wghp-obama-schools-speech-090903,0,1505064.story>)

6. This explains why our examples deal with noun phrases introduced by definite or zero articles.

- [19] Martin Townsend, editor of the Sunday Express, has made a personal appeal for her safe return, and said the paper had given its full support to her decision to enter the country illegally.  
(*The Guardian Europe*, 10/01/01)
- [20] Edgar Griffin, the father of the BNP leader Nick Griffin, was sacked as a vice-president of the Duncan Smith campaign in Wales after he admitted answering a BNP telephone inquiry line.  
(*The Guardian Europe*, 10/01/01)

18 Rephrasing the second unit of the appositive with an ARC is sometimes acceptable, sometimes clearly unacceptable:

- [14a] #Some expected Barack Obama, who is the president of the United States, to appoint a completely new economic team so as to implement another New Deal.
- [15a] #/?Bill Clinton, who is the former president of the United States, will attend an international seminar on AIDS and SARS (severe acute respiratory syndrome) and deliver a lecture on global AIDS prevention and control efforts, a seminar official said Friday.
- [16a] #/?Angela Merkel, who is the German chancellor, on Friday described Barack Obama's presidency as a "unique opportunity" to revive the Middle East peace process as the US leader continued his international tour with a stop in the historic eastern city of Dresden.
- [17a] ?Nancy Pelosi, who is the Speaker of the House, is among those on the Left now seeking to find common ground with the conservative populism that is sweeping across the United States.
- [18a] ?According to Arne Duncan, who is the secretary of education, the president will discuss the importance of hard work, educational goals and other topics.
- [19a] Martin Townsend, who is the editor of the Sunday Express, has made a personal appeal for her safe return, and said the paper had given its full support to her decision to enter the country illegally.
- [20a] Edgar Griffin, who is the father of the BNP leader Nick Griffin, was sacked as a vice-president of the Duncan Smith campaign in Wales after he admitted answering a BNP telephone inquiry line.

19 Such grammatical judgments on the pragmatic acceptability of [14a]–[20a] are not random and seem to be related to the (assumed) hearer new/old status of the identity relations:

- [14b] Barack Obama is the President of the United States.
- [15b] Bill Clinton is the former President of the United States.
- [16b] Angela Merkel is the German Chancellor.

[17b] Nancy Pelosi is the Speaker of the House (in Congress).

[18b] Arne Duncan is the Secretary of Education.

[19b] Martin Townsend is the editor of the *Sunday Express*.

[20b] Edgar Griffin is the father of the BNP leader Nick Griffin.

20 While [14b]/[15b] represent hearer old information for (probably) all addressees, [19b]/[20b] represent hearer new information, with [16b]/[17b]/[18b] representing intermediate cases (indeterminable hearer status) for English-speaking addressees ([16b] being more likely to be known than [17b] and [18b]). What the data show is that the more familiar the relation A is B, the less felicitous the use of an ARC, which makes the relation explicit (anaphoric pronoun + *be*), instead of a nominal appositive that keeps the identity relation implicit. This is what we call the familiarity degree constraint or "fame effect": the hearer old status of a relation prevents the use of an ARC.

21 But the problem for the definition of such a constraint is to show the existence of a link between assumed familiarity and the kind of structures that can be used in discourse. If one wants to go beyond sheer intuition, one needs a more 'scientific' confirmation. This is what we deal with in the next section.

### 3. How to define and establish the familiarity constraint

#### 3.1. Corpus investigation

##### 3.1.1. *The limits of traditional electronic corpora*

22 As a link exists between, on the one hand, the familiarity of an identity relation A is B (*e. g.* Barack Obama is the President of the US) and the (im)possibility of using specific structures such as the ones defined in the previous section, one must find a way to determine which structures can and which structures cannot be used to express this specific identity relation. The best way to determine which structures are pragmatically felicitous for a specific relationship is certainly a corpus search and the compiling of results that can be compared.

23 However, in this instance, the kind of corpus that can be used is a problem. Traditional corpora such as the ICE-GB corpus, which contains 1,000,000 words of written and spoken British English from the 1990s (<http://www.ucl.ac.uk/english-usage/projects/ice-gb/>), or even the British National Corpus (BNC), which contains 100,000,000 words of British English from the latter part of the 20<sup>th</sup> century (<http://www.natcorp.ox.ac.uk/>), do not suffice for the kind of data exploration we aim to perform here. Although very useful for providing attested examples of a specific word or construction, traditional corpora nevertheless reach their limits insofar as determining the frequency of a word or a construction in the overall language is concerned. The majority of words in the English lexicon occur less

than fifty times in the BNC (Kilgariff & Grefenstette, 2003: 4), which makes it impossible to compile figures or statistics concerning their use in the language. Some words or constructions may not even appear at all: for instance, Pomikálek *et al.* (2009) show that words like *hector* (V), *hedge* (N) or *heebie-jeebies* do not appear in the Susanne corpus, a 130,000-word corpus (<http://www.grsampson.net/RSue.html>) while they appear 37 times, 1,525 times and 0 time respectively in the BNC, which leads them to conclude that “corpora like Susanne are not usable for lexical knowledge, BNC is good for high frequency words but scarcely provides enough information to make informed generalizations on *hector* (verb) and certainly not for *heebie-jeebies*” (Pomikálek *et al.*, 2009: 4). As far as our project is concerned, Table 1 shows the number of occurrences that can be collected from the ICE-GB corpus and the BNC concerning the following queries about the most prominent political figures of the late 20<sup>th</sup> century – *i. e.* the period when most of the texts from these two corpora were collected – in the United Kingdom, *viz* Margaret Thatcher, John Major, and Tony Blair:

	ICE-GB	BNC
<i>Margaret Thatcher</i>	23	559
Prime Minister/PM Margaret Thatcher	1	48
Margaret Thatcher, (the) Prime Minister/PM	0	0
Margaret Thatcher, who is/was (the) PM of Britain/the United Kingdom	0	0
<i>John Major</i>	72	1453
Prime Minister/PM John Major	7	128
John Major, (the) Prime Minister/PM	0	0
John Major, who is/was (the) PM of Britain/the United Kingdom	0	0
<i>Tony Blair</i>	2	86
Prime Minister/PM Tony Blair	0	0
Tony Blair, (the) Prime Minister/PM	0	0
Tony Blair, who is/was (the) PM of Britain/the United Kingdom	0	0

Table 1. Number of hits in ICE-GB and BNC corpora

24

Although more information is required for us to draw valid conclusions, in particular regarding the dates of the texts (most of the texts from the ICE-GB corpus or the BNC date back to a period when Tony Blair had not yet been Prime Minister), what these figures do reveal is that the number of hits for such queries is clearly insufficient. Furthermore, most corpora are restricted in the geographical heterogeneity of speakers/writers and addressees: the ICE-GB and BNC corpora correspond to texts of British origin, for instance. To collect more examples from a wider variety of sources and see how they can be compared, we needed to resort to a (much) larger corpus.

### 3.1.2. *The web as corpus*

25 Like other researchers, we thus resorted to a tentative use of what some linguists consider the largest corpus available on earth, *viz* the internet, to try to determine whether the different structures used to express an identity relation show a high number of hits or not, with famous and 'not-so-famous' people. This is clearly experimental and the results provided in the article thus need to be accepted with caution.

26 Using the web as a corpus has been both advocated and rejected by linguists. While some researchers claim that the web is a gold mine in that it is the largest existing corpus, instantly available, easy to use, and provides an instant image of language use, others simply deny the fact that the web can be considered a corpus at all. The main objections to the use of the web as a corpus are the absence of tagging/parsing, its anarchic organization, its language errors (absence of editing or control over the grammaticality of the utterances), and its absence of exploitation by computational linguistics (Kilgarriff & Grefenstette, 2003). Also, while English is the dominant language on the web (85% of publicly-accessible webpages were in English in 2000; 72% in 2002 and 66% in 2005, according to Fletcher (2007), who compiled his results with other web specialists), only 35% of internet users were Anglophones in 2004, which means that not all webpages written in English are written by Anglophones themselves. As emphasized by Fletcher (2007: 37), "as the lingua franca of the digital frontier, English is both the target and source of contamination". On the other hand, advocates of the web as corpus speak highly of its unlimited size (according to Gulli & Signorini, 2005, the number of indexable webpages in 2005 was 11.5 billion), its free and easy access devoid of any copyright issues and its constant evolution both quantitatively and qualitatively.

27 Bearing in mind the many restrictions mentioned above, we did use the web as a corpus to compensate for the limits of traditional corpora. As our aim was to collect general estimates of frequencies for specific constructions, we used the hit counts provided by the search engine Google. Once again, this requires caution, as search engines provide a number of hits related to the number of pages, not the number of occurrences for the query, and can include several copies of the same webpage; in addition, the results show some arbitrariness, with results varying depending on the search engine that is used and also fluctuating from one day to the next (Kilgarriff, 2007: 148). Hit counts are also generally rough and provide only general estimates. All these restrictions naturally call for caution; however, the aim of our study was not to compile precise statistics for frequencies but to compare numbers of hits. To limit arbitrariness, we used the same search engine systematically and performed all searches on a specific person on the same day, within a few minutes<sup>7</sup>. We also restricted our searches to pages in English, which

7. Note that Fletcher (2007: 39) advocates the use of several search engines, even different regional versions of the same search engine, and at several weeks interval, with various orders for the queries, "to ensure stable counts and tolerable variance".

is of course no guarantee that all writers were native speakers of English. We thus used the web as corpus (WaC) as opposed to the web for corpus (WfC), according to Fletcher (2007)'s distinction, with a hunting approach, as opposed to grazing and browsing approaches (see Fletcher, 2007: 28-30)<sup>8</sup>.

- 28 We are not the only ones using the web as a corpus in such a way. An interesting study by Keller & Lapata (2003) has shown that, in order to overcome the “data sparseness” of edited corpora, counts obtained from the web are as legitimate as frequency counts obtained from traditional, edited, electronic corpora. Using the BNC and the North American News Text Corpus (NANTC) on the one hand and the internet through search engine searches performed with AltaVista and Google on the other hand, they compiled frequencies for predicate-argument bigrams, *i. e.* Noun-Noun, Adjective-Noun, Verb-Object combinations (*e. g. process user, hungry animal, fulfill obligation*), 540 in total. Their results show that counts obtained through web searches are highly correlated with counts obtained from both the BNC and the NANTC. This legitimizes the use of the web, in spite of its “noisiness” (*i. e.* the fact that it is not edited in the same way as a traditional corpus, that queries can provide false positives and also return webpages that do not even include the query), to determine the frequencies of both “seen” and “unseen” bigrams, *i. e.* bigrams that appear and do not appear in edited corpora respectively. In addition, they show that web frequencies can also predict human plausibility judgments. In the same vein, a study by Grefenstette (1998: 3) has shown that frequency counts retrieved from the web can determine the correct translation among *n* possible translations: *e. g.* while the French N<sub>1</sub> P N<sub>2</sub> compound *groupe de travail* has potentially fifteen different N<sub>2</sub> N<sub>1</sub> translations in English (*groupe* can be translated as *cluster, group, grouping, concern* and *collective*; *travail* can be translated as *work, labour* or *labor*), statistics compiled from the web show that *work group* is much more frequent than any other combination, indicating it is the “most likely domain-independent translation”.

### 3.1.3. Results

- 29 We made specific queries like “US President (Barack) Obama”/“American President (Barack) Obama” *vs* “(Barack) Obama, (the) President of the US/United States/America” on the search engine Google and compared the number of hits for the expression of an identity relation through the use of a nominal appositive (*e. g.* “Barack Obama, the President of the US”), an ARC (*e. g.* “Barack Obama, who is the President of the US”) or a pre-modifier (*e. g.* “President of the US Barack Obama”). All Google searches were performed in September and October 2009. Below are the results for Barack Obama:

8. Alternative options would be to use a local corpus, isolated from the web such as BiWeC (see Pomikálek *et al.*, 2009), cleaned of all irrelevant pages, or to use a software such as KWICFinder (see Fletcher, 2001, 2007) that allows for linguistics-oriented queries.



- [21] Barack Obama, who is (the) President of the United States/America/the US 3  
 American/United States/US President Barack Obama 125,155,000  
 Barack Obama, (the) President of the United States/America/the US 58,420,987

30 These results agree with the idea that a link might exist between the hearer old status of the relation (everyone must know in 2009 who Barack Obama is) and the impossibility to find an ARC. Interestingly, the results for the use of an ARC (3 hits only) show very specific contexts. Examples [22] and [23] below are both taken from situational contexts where the addressees are children (the third example is different and is discussed in footnote 9):

- [22] We were watching the inauguration of Barack Obama who is the President of the United States now.  
 (child's blog, <http://paytonandellie247.blogspot.com/feeds/posts/default>)
- [23] The notable alumni of Columbia include one of the most powerful men in the world: Barack Obama, who is the President of the United States, was a part of Columbia College Class of 1983.  
 (Pakistani magazine for the youth, <http://jang.com.pk/thenews/oct2009-weekly/us-16-10-2009/p22.htm>)

31 In the same way, it is interesting to note that we do find examples with ARCs to specify Barack Obama's position when he was president-elect ([24]–[26]), *i. e.* between his election in November 2008 and the inauguration day ceremony in January 2009. The specific president-elect status is not as famous or salient as the presidential status and the identity relation is also only recently established, so more hearer new.

- [24] As we congratulate the American people and more specifically, Senator Barrack Obama who is now President-elect, it is important to reflect on the journey that he has travelled so far.  
 (Kenyan parliamentary debate, 11/05/08, <http://www.thepostemail.com/2010/04/14/kenyan-parliament-proclaimed-obamas-birth-there-in-november-2008/>)
- [25] [Stanley Ann Dunham Soetoro] is the mother of Barack Hussein Obama, who is the President-elect of the United States.  
 (*Christopedia*, <http://christopedia.atwiki.com/>)
- [26] But ominously, Barak Obama, who is the president-elect of the US, is on record of having said that India must 'solve' the Kashmir issue to Pakistan's satisfaction.  
 (*Sify News*, Indian news website, <http://sify.com/news/>)

32 Such results reinforce the idea that with ARCs, the relation must not be familiar, but needs to be asserted explicitly, because of assumed unfamiliarity (children's blog or magazine), newness of the relation (president-elect examples)

or, even, geographical distance (the president-elect examples often appear on Indian and Kenyan websites)<sup>9</sup>.

33 The use of an ARC is then to be contrasted with the use of a nominal appositive or a pre-modifier, which both present the relation as given, familiar, unproblematic.

34 If we now consider other relations concerning perhaps less famous people, *i. e.* referents whose association with an identity relation is less salient, more hearer new for the addressees, such as William Taft or Nancy Pelosi, in opposition with Barack Obama and Angela Merkel for instance, here is what Google searches provide us with:

[27]	Angela Merkel, who is the Chancellor of Germany	6
	German Chancellor Angela Merkel	3,720,000
	Angela Merkel, the Chancellor of Germany	84,900
[28]	William (H(oward)) Taft, who was President	155
	Former (US) President William (H(oward)) Taft	82,974
	William (H(oward)) Taft, President	5,158
[29]	Nancy Pelosi, who is (the) Speaker of the House (of Representatives)	22,412
	Speaker of the House (of Representatives) Nancy Pelosi	168,900
	Nancy Pelosi, (the) Speaker of the House (of Representatives)	4,527

35 What these results show is that there is a clear difference between Barack Obama/Angela Merkel on the one hand and William Taft/Nancy Pelosi on the other for the possibility to find an ARC to express the following identity relations:

- [21a] Barack Obama is (the) President of the United States
- [27a] Angela Merkel is (the) Chancellor of Germany
- [28a] William (H(oward)) Taft was President

9. We have also shown that another possibility for the use of an ARC is the need to re-assert an identity relation whose legitimacy has been/can be questioned, as is shown by the following two examples (the first one is made up and taken from Loock (2010: 172) and the second one from an internet forum where the speaker interacts with an addressee questioning the legitimacy of B. Obama's being President of the US):

[i] Barack Obama is already being criticized for his economic policy to tackle the economic crisis. Some of his fiercest opponents even consider him to be incompetent as far as economic matters are concerned and even deny him his right to the title of president. These people keep reminding us that the people that really are in command of the country's economy are industrialists, who know what should be done to boost economic development. These industrialists are sometimes called the real presidents of the United States. But the president of the United States, who is Barack Obama, has repeatedly re-asserted that he is the one in charge of the country, no matter what his opponents might say.

[ii] No one gets to take their original birth certificate home with them, idiot. You can get a certified copy for identification from the county, or whatever government agency is responsible for maintaining this information, but you don't get to go anywhere with the original. You, me or Barack Obama, who is the President of the United States whether or not you've got the intelligence to grasp that (<http://thinkprogress.org>).

[29a] Nancy Pelosi is (the) Speaker of the House (of Representatives)

- 36 While using an ARC instead of a nominal appositive or a pre-modifier is exceptional for Barack Obama and Angela Merkel, who are both known internationally, the results show that it is not exceptional to find an ARC to express an identity relation for William Taft, probably one of the less famous among American presidents, and definitely frequent and natural for Nancy Pelosi, whose name might be familiar for an American addressee but whose function (Speaker of the House) is not that well-known.

### 3.2. Limits

- 37 However, the picture given above might be too simplistic. Results for Arne Duncan, the American Secretary of Education, for instance, are not that clear-cut:

[30]	Arne Duncan, who is (the) Secretary of Education	3
	Arne Duncan, (the) Secretary of Education	3,728
	Secretary of Education Arne Duncan	2,847,000

- 38 Even for an American addressee, the identity relation *Arne Duncan is the Secretary of Education* is not systematically hearer old. His name and function are less likely to be hearer old than the name and function of Nancy Pelosi, for instance. And yet, the number of hits for the use of an ARC is the same as for Barack Obama. Making the identity relation explicit through the use of a relative pronoun and the verb *be* is not the privileged way of referring to the American Secretary of Education.

- 39 It is also important to note that the familiarity constraint is only a one-way constraint: the "fame effect" provides an explanation as to why an ARC is infelicitous when the identity relation is too salient/hearer old for the addressees. However, this does not mean that non salient/non familiar identity relations do require the use of an ARC. Consider the following nominal appositives [31]–[33] alongside pre-modifiers [31a]–[33a], which are all perfectly grammatical and pragmatically felicitous:

[31] Mr Miliband is scheduled to hold bilateral talks with Mr A. Raja, the Indian Environment Minister.

(adapted from <http://ukinindia.fco.gov.uk/en/about-us/working-with-india/ministerial-visits/visits-in-2007>)

[31a] Mr Miliband is scheduled to hold bilateral talks with Indian Environment Minister Mr A. Raja.

[32] AT&T, the No. 1 long-distance carrier, will announce Thursday plans to sell Internet phone service to consumers and to more businesses.

(*USA Today*, 11/12/03)

[32a] No. 1 long-distance carrier AT&T will announce Thursday plans to sell Internet phone service to consumers and to more businesses.

[33] Hassan Rohani, the chief Iranian negotiator, said after three hours of talks in Geneva yesterday that the Europeans had until the end of July to come up with a more concrete offer.

(<http://www.buzzle.com/editorials/5-25-2005-70530.asp>)

[33a] Chief Iranian negotiator Hassan Rohani said after three hours of talks in Geneva yesterday that the Europeans had until the end of July to come up with a more concrete offer.

40 In each of these examples, the use of nominal appositives or pre-modifiers expresses an identity relation, an identity relation that is hearer new for most of the addressees:

[31b] Mr A. Raja is the Indian Environment Minister.

[32b] AT&T is the No. 1 long-distance carrier.

[33b] Hassan Rohani is the chief Iranian negotiator.

41 The difference with examples that include Barack Obama and Angela Merkel is that the use of ARCs is possible:

[31c] Mr Miliband is scheduled to hold bilateral talks with Mr A. Raja, who is the Indian Environment Minister.

[32c] AT&T, which is the No. 1 long-distance carrier, will announce Thursday plans to sell Internet phone service to consumers and to more businesses.

[33c] Hassan Rohani, who is the chief Iranian negotiator, said after three hours of talks in Geneva yesterday that the Europeans had until the end of July to come up with a more concrete offer.

42 When the identity relation is hearer new, nominal appositives, pre-modifiers and ARCs are possible. The familiarity constraint is thus a one-way constraint: familiarity excludes the use of an ARC but non familiarity does not require the use of an ARC.

### 3.3. Another illustration of the "fame effect": nominal appositives vs pre-modifiers

43 In the previous subsection, non-restrictive, relevance-oriented pre-modifiers have been presented as equivalent to nominal appositives as expressing an identity relationship that can be either hearer new or hearer old. But once again, it seems that the degree of familiarity has an influence on the felicitousness of the use of a nominal appositive or a pre-modifier. Consider the following examples, which mention Jaycee Dugard, an American young woman who was abducted and kept prisoner for twenty-eight years before being released in late August 2009 (all examples are

extracted from different news websites such as [www.bbc.co.uk](http://www.bbc.co.uk), [www.theguardian.co.uk](http://www.theguardian.co.uk); in brackets the posting dates are mentioned)<sup>10</sup>:

- [34] US kidnap victim Jaycee Dugard speaks publicly for the first time since being found living with her alleged abductors for 18 years.  
(11/14/09)
- [35] The newspapers have fresh revelations and theories on kidnap victim Jaycee Lee Dugard.  
(09/01/09)
- [36] Kidnapped US woman Jaycee Lee Dugard bonded with her alleged captor and helped to run his printing firm, according to reports.  
(08/31/09)
- [37] The story of the Californian kidnap victim, Jaycee Lee Dugard, again generates wide coverage in all the newspapers.  
(08/30/09)
- [38] Widespread coverage of the story of Jaycee Lee Dugard – the US girl imprisoned for 18 years – draws comparisons to Natascha Kampusch in Austria.  
(08/29/09)

44 What we notice is that pre-modifiers are used in more recent press articles than nominal appositives. This suggests that the identity relation *Jaycee Lee Dugard is a Californian kidnap victim imprisoned for 18 years* is more salient, more hearer old when a pre-modifier is used than when a nominal appositive is used.

45 If we try and corroborate this intuition with figures compiled through Google searches for nominal appositives following proper nouns (Barack Obama, (the) President of the US) *vs* premodifiers (US President Barack Obama), here are the results that we obtain (searches also performed in September–October 2009):

	Nominal appositive	Pre-modifier
Barack Obama	58,420,987	125,155,000
Angela Merkel	84,900	3,720,000
Arne Duncan	3,728	2,847,000
Nancy Pelosi	4,527	168,900
Martin Townsend	7,920	4,830
Carolyn Williams	4,467	3,111

Table 2. Number of hits nominal appositives *vs* pre-modifiers

10. The dates only cover a very short time-span (August 29–November 14), but this corresponds to the period when the event of Jaycee Lee Dugard's release was covered by the media. The covered time-span on one unique example prevents from drawing any firm conclusions, but it is interesting to note the evolution concerning this particular case.

46 What these figures suggest is that the more familiar the A is B relation, the more a pre-modifier is used: for Barack Obama, Angela Merkel, Arne Duncan, and Nancy Pelosi, pre-modifiers is the favored syntactic structure whereas for Martin Townsend (*Sunday Express* editor) and Carolyn Williams (an American literature professor), who are completely unknown to the general public, nominal appositives are favored over pre-modifiers. Given the limits of the corpus, more investigation is required concerning this question to confirm this tendency and reach firm conclusions.

#### 4. Conclusion

47 The aim of this paper was to show that the choice between the different syntactic means to express a relevance-oriented non-restrictive identity relation is not random. In particular, our aim was to show that among the constraints governing writers' choices between allostructures such as ARCs, nominal appositives, pre-modifiers... is the familiarity degree constraint or "fame effect": writers' choices between n allostructures are dependent on the assumed knowledge of the addressees. For instance, the use of an ARC is only felicitous when the relation is assumed to be hearer new for the addressees or needs to be reasserted because its legitimacy has been questioned, whereas the use of a nominal appositive is felicitous independently of the hearer new/hearer old informational status of the relation. The use of non-restrictive pre-modifiers is favored when the relation has already been established and is therefore assumed to be salient for the addressees.

48 Such results naturally strengthen the idea that the grammatical form of a sentence is clearly dependent on the situational context (identity of speakers/writers and addressees) of the whole text. When there are several possibilities for packaging the same informational content, the same logico-semantic meaning, the identity of the addressees and, above all, the way the latter are perceived by the speaker/writer have an influence on his/her grammatical choices.

49 Still, many questions remain unanswered. Other methods for exploiting the web should be investigated, such as Fletcher's application KWiCFinder or Pomikálek *et al.* (2009)'s method for creating off-line corpora such as BiWeC, to corroborate our results. Moreover, a way to evaluate the "fame" of a specific person and his/her official function/role/occupation needs to be found: while it is easy to determine that the relation *Barack Obama is President of the United States* is hearer old, in non-technical terms famous, this is much more difficult for relations like *Angela Merkel is the German Chancellor* or *Nancy Pelosi is Speaker of the House*. The informational status (hearer old/new) for A is B relations could be demonstrated through psycholinguistic experiments that calculate the number of seconds necessary to establish a link between A and B (*vs* C or D for instance). All this is naturally left open for future research to refine the definition and role of what we call the "fame effect".

## References

- BLANCHE-BENVENISTE, C. & CHERVEL, A. 1966. Recherches sur le syntagme substantif. *Cahiers de lexicologie* 9: 3-37.
- BURTON-ROBERTS, N. 1975. Nominal Apposition. *Foundations of Language* 13: 391-419.
- CHAFE, W.L. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics and Point of View. In C. LI (ed.), *Subject and Topic*. New York: Academic Press: 25-55.
- DANEŠ, F. 1966. A Three-Level Approach to Syntax. In F. DANEŠ *et al.* (eds), *Travaux linguistiques de Prague*. Alabama: University of Alabama Press. Vol. 1: 225-240.
- FLETCHER, W.H. 2001. Concordancing the Web with KWICFinder. *Proceedings of the American Association for Applied Corpus Linguistics. Third North American Symposium on Corpus Linguistics and Language Teaching*. Available online from <http://www.kwicfinder.com>.
- FLETCHER, W.H. 2007. Concordancing the Web: Promise and Problems, Tools and Techniques. In M. HUNDT, N. NESSELHAUF & C. BIEWER (eds), *Corpus Linguistics and the Web*. Amsterdam: Rodopi: 25-45.
- GREFENSTETTE, G. 1998. The World Wide Web as a Resource for Example-Based Machine Translation Tasks. *Proceedings of Aslib Conference on Translation and the Computer* 21. Retrieved on [http://www.xrce.xerox.com/content/download/16526/118269/file/gg\\_aslib.pdf](http://www.xrce.xerox.com/content/download/16526/118269/file/gg_aslib.pdf).
- GULLI, A. & SIGNORINI, A. 2005. The Indexable Web is More than 11.5 Billion Pages. *Proceedings of 14<sup>th</sup> International World Wide Web Conference, Chiba, Japan*. Retrieved on [http://www.di.unipi.it/~gulli/papers/f692\\_gulli\\_signorini.pdf](http://www.di.unipi.it/~gulli/papers/f692_gulli_signorini.pdf).
- HIGGINS, R.F. 1979. *The Pseudo-cleft Construction in English*. New York: Garland Publishing.
- KELLER, F. & LAPATA, M. 2003. Using the Web to Obtain Frequencies for Unseen Bigrams. *Computational Linguistics* 29 (3): 459-484.
- KILGARRIFF, A. 2007. Googleology is Bad Science. *Computational Linguistics* 33 (1): 147-151.
- KILGARRIFF, A. & GREFENSTETTE, G. 2003. Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics* 29 (3): 333-347.
- LAMBRECHT, K. 1996. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- LOOCK, R. 2003. Les fonctions discursives des propositions subordonnées relatives appositives en discours. *Anglophonia* 12: 113-131.
- LOOCK, R. 2005. *La proposition subordonnée relative appositive à l'écrit et à l'oral en anglais contemporain : fonctions discursives et structures concurrentes*. Thèse de Doctorat. Université Charles de Gaulle, Lille.
- LOOCK, R. 2007. Appositive Relative Clauses and their Functions in Discourse. *Journal of Pragmatics* 39: 336-362.
- LOOCK, R. 2010. *Appositive Relative Clauses in English: Discourse Functions and Competing Structures*. Studies in Discourse and Grammar 22. Amsterdam – Philadelphia: John Benjamins.
- MIKKELSEN, L. 2004. Reexamining Higgins' Taxonomy: a Split in the Identificational Class. Paper presented at the 78<sup>th</sup> Linguistics Society of America meeting (Boston, January).

- POMIKÁLEK, J., RYCHLY, P. & KILGARRIFF, A. 2009. Scaling to Billion-plus Word Corpora. *Advances in Computational Linguistics* 41: 3-13.
- PRINCE, E.F. 1981. Toward a Taxonomy of Given/New Information. In P. COLE (ed.), *Radical Pragmatics*. New York: Academic Press: 223-254.
- PRINCE, E.F. 1992. The ZPG Letter: Subjects, Definiteness, and Information-Status. In S. THOMPSON & W. MANN (eds), *Discourse Description: Diverse Analyses of a Fund Raising Text*. Amsterdam – Philadelphia: John Benjamins: 295-325.
- SPERBER, D. & WILSON, D. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.
- VALLDUVÍ, E. 1992. *The Informational Component*. PhD thesis. University of Pennsylvania.
- VALLDUVÍ, E. 1993. *Information packaging: a survey*. Report prepared for WOPIS (Word Order, Prosody, and Information Structure). University of Edinburgh.